

**Notice of Allowability**

Application No.

09/941,301

Applicant(s)

BAKIS ET AL.

Examiner

James S. Wozniak

Art Unit

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 9/6/2005.
2. ☒ The allowed claim(s) is/are 1-4, 7, 10-14, 19, 21, 23-28, 30, 31 and 35-43 (now 1-29).
3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some\* c) ☐ None of the:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

\* Certified copies not received: \_\_\_\_\_.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.

**THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
- (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
- 1) ☐ hereto or 2) ☐ to Paper No./Mail Date \_\_\_\_\_.
- (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date \_\_\_\_\_.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

**Attachment(s)**

1. ☒ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. ☐ Information Disclosure Statements (PTO-1449 or PTO/SB/08), Paper No./Mail Date \_\_\_\_\_
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application (PTO-152)
6. ☐ Interview Summary (PTO-413), Paper No./Mail Date \_\_\_\_\_
7. ☒ Examiner's Amendment/Comment
8. ☒ Examiner's Statement of Reasons for Allowance
9. ☐ Other \_\_\_\_\_

W. R. YOUNG  
PRIMARY EXAMINER

## **DETAILED ACTION**

### ***Response to Amendment***

1. In response to the office action from 6/21/2005, the applicant has submitted a request for continued examination, filed 9/6/2005, amending claims 1, 5, 7, 10, 11, 19, 21, 25, 35, and 43, while canceling claims 8-9, 15-18, 22, 32-34 and 44 and arguing to traverse the art rejection based on the limitation regarding amended limitations (*Amendment, pages 17-19*). The applicant's arguments have been fully considered and Claims 1-4, 7, 10-14, 19, 21, 23-28, 30-31, and 35-43 are allowable over the prior art of record for the reasons given below and with respect to the examiner's amendment.

## **EXAMINER'S AMENDMENT**

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it **MUST** be submitted no later than the payment of the issue fee.

3. Authorization for this examiner's amendment was given in a telephone interview with Richard Hinson (Reg. No. 47,652) on 10/17/2005.

4. The application has been amended as follows:

Amend the claims as follows:

1. In a text-to-speech system, a method of converting text-to-speech comprising:  
receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

generating processed input by parsing and normalizing said text input;

comparing said processed input to at least one entry in a text-to-speech cache memory, wherein said entry in said text-to-speech cache memory specifies a corresponding spoken output, wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information, and wherein each spoken output has an assigned score;

if said processed input matches one of said entries in said text-to-speech cache memory, providing said spoken output specified by said matching entry and rendering said spoken output according to said plurality of attributes associated with said text input;

if said processed input fails to match one of said entries, generating an additional spoken output with a text-to-speech engine, generating an entry that specifies said additional spoken output, assigning a score to said additional spoken output, storing said additional spoken output and assigned score in said cache memory, and rendering said spoken output with the text-to-speech engine according to said plurality of attributes associated with said text input, wherein each assigned score is an updatable score computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score;

if the cache memory is full when said additional spoken output is generated, deleting from said cache memory a spoken output having a lower score; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of corresponding spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

Art Unit: 2655

2. The method of claim 1, wherein said text-to-speech cache entries include an intermediate output which is not a digitally encoded audio file; and wherein said text-to-speech engine converts said intermediate output to said spoken output.
3. The method of claim 1, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.
4. The method of claim 1, further comprising logging each said match of said text input with a text-to-speech cache entry.
5. (Cancelled)
6. (Cancelled)
7. The method of claim 1, further comprising periodically updating each said score.
8. (Cancelled)
9. (Cancelled)
10. The method of claim 1, further comprising comparing said attributes of said received text input with attributes of said entries in said text-to-speech cache memory.
11. A method of converting text-to-speech using a text-to-speech cache memory having a plurality of entries, wherein said entries comprise a processed form specifying a spoken output, wherein said processed form specifying spoken output does not comprise a digitally encoded audio file, said method comprising:

receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

assigning a score to each of said entries, wherein each score is an updatable score computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score;

if said text input matches one of said entries in said text-to-speech cache memory, providing said processed form specified by said matching entry to a text-to-speech engine;

said text-to-speech engine converting said processed form to said spoken output and rendering said spoken output according to said plurality of attributes associated with said text input; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

12. The method of claim 11, wherein the determined form of said text input comprises at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine.

13. The method of claim 11, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

14. The method of claim 11, further comprising logging each said match of said text input with a text-to-speech cache entry.

15. (Cancelled)

16. (Cancelled)

17. (Cancelled)

18. (Cancelled)

19. A method of converting text-to-speech comprising:

storing a plurality of entries in a text-to-speech cache memory, wherein the text-to-speech cache memory is directly and locally coupled to at least one text-to-speech engine, wherein each said entry comprises a processed form specifying a spoken output, and wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information;

assigning a score to each one of said plurality of entries;

receiving a text input;

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

when at least one of the plurality of entries in said text-to-speech cache memory is matched to said determined form, retrieving the processed form for the matching entry from the text-to-speech cache memory, and using the processed form to generate said spoken output based on said attributes;

when at least one of the plurality of entries in said text-to-speech cache memory is not matched to said determined form, using the at least one text-to-speech engine to generate said spoken output;

logging when one of said plurality of entries in said text-to-speech cache memory is matched to said received text input

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory; and

periodically updating said score for each one of said plurality of entries of said text-to-speech cache memory, wherein an updated score is computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score.

20. (Cancelled)

21. A text-to-speech system comprising:

a text-to-speech engine for receiving text inputs and a plurality of attributes associated with said text and for producing a spoken output representative of said received text, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input; and

a text-to-speech cache memory for storing selected entries corresponding to received text inputs and a score assigned to each entry, wherein said entries specify spoken outputs corresponding to said selected received text inputs, wherein at least one processing interaction occurs between the speech-to-text engine and the text-to-speech cache memory when the text-to-speech engine uses the text-to-speech memory cache to generate the spoken output responsive to receiving text, said processing interactions comprising at least one interaction selected from the group consisting of a pre-processing interaction where the received text is processed into an intermediate form before being compared to entries of the text-to-speech cache that are stored in said intermediate form and a post-matching interaction where the specified spoken outputs retrieved from the text-to-speech cache memory are processed by the text-to-speech engine to generate the spoken output according to the associated attributes, and wherein each score is an updatable score computed by multiplying a previous score times a constant between zero and one

Art Unit: 2655

and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score.

22. (Cancelled)

23. The text-to-speech system of claim 21, wherein said text-to-speech cache entries include said spoken output, and wherein the processing interaction is a pre-processing interaction, and wherein the intermediate form comprises normalized text that represents a standardized version of the text input.

24. The text-to-speech system of claim 21, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

25. A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

generating processed input by parsing and normalizing said text input;

comparing said processed input to at least one entry in a text-to-speech cache memory, wherein said entry in said text-to-speech cache memory specifies a corresponding spoken output, wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and a score corresponding to each entry, ~~callback information, and~~ wherein each spoken output has an ordinal ranking and wherein each score is an updatable score computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score;



if said processed input matches one of said entries in said text-to-speech cache memory, providing said spoken output specified by said matching entry and rendering said spoken output according to said plurality of attributes associated with said text input;

if said processed input fails to match one of said entries, generating an additional spoken output with a text-to-speech engine, generating an entry that specifies said additional spoken output, assigning an ordinal ranking to said additional spoken output, storing said additional spoken output and assigned ordinal ranking in said cache memory, and rendering said spoken output with the text-to-speech engine according to said plurality of attributes associated with said text input;

if the cache memory is full when said additional spoken output is generated, deleting from said cache memory a spoken output having a lower ordinal ranking; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of corresponding spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

26. The machine-readable storage of claim 25, wherein said text-to-speech cache entries include an intermediate output which is not a digitally encoded audio file; and wherein said text-to-speech engine converts said intermediate output to said spoken output.

27. The machine-readable storage of claim 25, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

28. The machine-readable storage of claim 25, further comprising logging each said match of said text input with a text-to-speech cache entry.

29. (Cancelled)

Art Unit: 2655

30. The machine-readable storage of claim 25, further comprising removing one of said entries in said text-to-speech cache memory.

31. The machine-readable storage of claim 25, wherein each said entry in said text-to-speech cache memory has a score, said machine-readable storage further comprising periodically updating each said score.

32. (Cancelled)

33. (Cancelled)

34. (Cancelled)

35. A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

storing a plurality of entries in a text-to-speech cache memory, wherein each one of said entries comprises a processed form specifying a spoken output wherein said processed form specifying spoken output does not comprise a digitally encoded audio file;

assigning a score to each one of said plurality of entries, each score being an updatable score computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score;

receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

if said text input matches one of said entries in said text-to-speech cache memory, providing said processed form specified by said matching entry to a text-to-speech engine; said text-to-speech engine converting said processed form to said spoken output and rendering said spoken output according to said plurality of attributes associated with said text input; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

36. The machine-readable storage of claim 35, wherein the determined form of said text input comprises at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine.

37. The machine-readable storage of claim 35, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

38. The machine-readable storage of claim 35, further comprising logging each said match of said text input with a text-to-speech cache entry.

39. The machine-readable storage of claim 35, wherein said text input does not match an entry in said text-to-speech cache memory, said method further comprising:

determining a spoken output corresponding to said text input by using the text-to-speech engine to text-to-speech convert the text input; and

storing an entry in said text-to-speech cache memory corresponding to said text input, wherein said entry specifies said determined spoken output.

40. The machine-readable storage of claim 35, further comprising removing one of said entries in said text-to-speech cache memory.

Art Unit: 2655

41. The machine-readable storage of claim 35, wherein each said entry in said text-to-speech cache memory has a score, said machine-readable storage further comprising periodically updating each said score.

42. The machine-readable storage of claim 41, further comprising removing one of said entries in said text-to-speech cache memory having a lowest score.

43. A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

storing a plurality of entries in a text-to-speech cache memory, wherein the text-to-speech cache memory is directly and locally coupled to at least one text-to-speech engine, wherein each said entry comprises a processed form specifying a spoken output, and wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information;

assigning a score to each one of said plurality of entries;

receiving a text input;

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

when at least one of the plurality of entries in said text-to-speech cache memory is matched to said determined form, retrieving the processed form for the matching entry from the text-to-speech cache memory, and using the processed form to generate said spoken output based on said attributes;

when at least one of the plurality of entries in said text-to-speech cache memory is not matched to said determined form, using the at least one text-to-speech engine to generate said spoken output;

logging when one of said plurality of entries in said text-to-speech cache memory is matched to said received text input

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory; and

periodically updating said score for each one of said plurality of entries of said text-to-speech cache memory, wherein an updated score is computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score.

44. (Cancelled)

#### ***Allowable Subject Matter***

5. **Claims 1-4, 7, 10-14, 19, 21, 23-28, 30-31, and 35-43** are allowable over the prior art of record.

6. The following is an examiner's statement of reasons for allowance:

With respect to **Claims 1, 11, 19, 21, 25, and 35**, the prior art of record fails to specifically teach or fairly suggest a text-to-speech caching method, system, or computer readable medium containing a program that matches a parsed input text to an existing spoken output entries in a text-to-speech cache and provides the spoken output if a match is found. If a match is not found, the text-to-speech caching system generates and stores a new spoken output entry through the use of a text-to-speech engine and assigns a score to said entry that is updated by multiplying a previous scores times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last update. The prior art of record also fails to specifically teach or fairly suggest the aforementioned claim

limitations in combination with a display that highlights words in coordination with a speech output. Although Richard et al (*U.S. Patent: 5,924,068*) teaches a temporary memory that stores entries specifying spoken outputs (pronunciation data) and creates and stores a new entity when an input text match is not found, Richard et al does not teach the aforementioned display feature nor the cache management score updating feature of the presently claimed invention. Carter et al (*U.S. Patent: 6,600,814*) teaches a cache storing previously converted speech entities, which also utilizes a recency of access algorithm, but does not explicitly teach how the algorithm is utilized or updated and is silent with respect to the highlighting feature of the presently claimed invention. Thus, Claim 1 is allowable over the prior art of record.

The remaining dependent claims further limit allowed independent claims, and thus, are also allowable over the prior art of record.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

### ***Conclusion***

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

Luther (*U.S. Patent: 5,500,919*)- teaches a text-to-speech synthesizer that highlights spoken words, but does not teach the use of a cache memory or the cache memory updating feature of the presently claimed invention.

Kiraly et al (*U.S. Patent: 6,324,511*)- teaches a text-to-speech system that sequentially highlights text as the corresponding words are spoken, but does not teach the use of a cache memory or the cache memory updating feature of the presently claimed invention.

Walker et al (*U.S. Patent Pub: 2001/0048736*)- teaches a cache memory that stores speech data that has undergone text-to-speech processing, but does not teach cache entity scores or a highlighting capability.

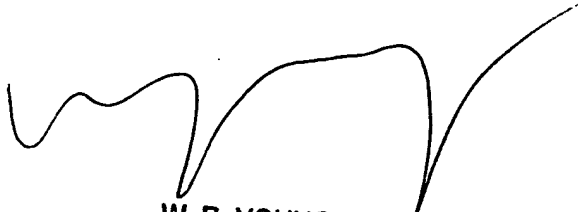
Guedalia et al (*U.S. Patent Pub: 2002/0091524*)- teaches a cache memory located within a text-to-speech server that stores previously speech converted text phrases, but does not teach cache entity scores or a highlighting capability.

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (571) 272-7632. The examiner can normally be reached on M-Th, 7:30-5:00, F, 7:30-4, Off Alternate Fridays.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Wayne Young can be reached on (571) 272-7582. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James S. Wozniak  
10/20/2005



W. R. YOUNG  
PRIMARY EXAMINER